Universal Networking Language (UNL)

a Means to Bridge the Digital Divide

Dr. Saqer Abdel-Rahim
Dr. Asad Abu Libdeh
Fakher Sawalha
Mohammed K. Odeh

Computer Technology Training and Industrial Studies Center
Royal Scientific Society
March 19, 2002
Introduction

The last decade of the 20th century witnessed an unimaginary acceleration in the development of information technology in all fields of life. The decade also witnessed a great increase in the spread and popularity of the internet. Tens of millions of people of almost all levels of education and attitudes, of different jobs all over the world use the internet, and for different purposes. English is the main language of the internet. Understandably, not all people know English. Urgent need, therefore, arose to develop interlanguage translation programs. Among those who did their best to tackle this problem was the United Nations University/ Institute of Advanced Studies (UNU/IAS). The Institute conducted a review of all internationally available machine translation programs, and as a result, decided to start devising a better, more efficient and more workable technique via the internet; a technique more accessible to the users of the internet (whose native languages are of course different) and which makes shorter the long track of bilingual translation programs, each of which needs a special software. The baby was the Universal Networking Language (UNL). The main aim of the UNL is to overcome language barriers.

The UNL is based on developing an intermediary language system whereby any written text can be converted to many languages (all languages involved in the program) and simultaneously, all texts written in different languages can be converted to that particular language.

The UNL is a conceptual computer semantic language, in that words which express the same concept are grouped together under one main category. It is system for multi-lingual communications.

Not only will the UNL provide people with equal opportunities to participate in economic and social activities but it will also mean that all people can access science, technology, economic and social information in their own language. People throughout the world will be provided with equal opportunity in business and education of science and technology. In particular, the fair distribution of information will support the activities of international organizations such as the United Nation and make them more truly international.
It is worth mentioning that UNU/IAS created the UNDL program which is now carried out by the non-profit UNDL Foundation. The Foundation was established at the beginning of the year 2001 with its headquarters in Geneva, Switzerland; as one of the UN organizations. It was officially registered three months later.

By way of illustration, the following figure shows the UNL system and how one language is converted to another:

In this paper, we highlight the most important features of this program, its development, the languages and countries involved, and what has been accomplished so far by this program.

**UNL solves Information Monopolies**

Most of the information in the world is written in the author’s mother tongue. There is little sharing of information written in different languages and this information is used only by locals. Information for the whole world is often written in English but, for people whose mother tongue is not English, it is often difficult to access this information. Conversely, information that is not written in English is difficult to understand for people for whom that language is not their mother tongue.

Even now in networks there is a flood of information written in a number of languages but the language barrier is hindering the access of this information. Information written in the transparent UNL can,
via a network, be received by anyone in their mother tongue. Once information written in one language is “enconverted” into UNL it will be able to be shared by anyone in the world.

The use of UNL will be of a great benefit in terms of economy in that it save time and money.

UNL also needs no language analysis. These two points alone give the UNL the edge over other ways of translation.

However, the UNL is not necessarily accessible to all people. It is made up of logical description format that represents the meaning of sentences and so, at times, can be understood only by trained specialists. To solve this, an editing system is provided so that documents can be entered in UNL. A number of editing systems could be possible that would meet the needs of particular uses. For native users who enter their own information, there will probably be UNL editing systems where information can be entered in language similar to their native tongue and there will probably be other editing systems for experts that enter large quantities of information.

We can also expect the advent of a new occupation, the UNL writer, who creates information in UNL. A UNL writer will be different from the translator of the past and will convert information from any language into UNL. That is, a UNL writer will not be translating one language into another, but will be building a bridge across to all other languages.

Currently, the UNL includes 16 languages. These include the six official languages of the United Nations (Arabic, Chinese, English, French, Russian and Spanish), in addition to ten other widely spoken languages (German, Hindi, Italian, Indonesian, Japanese, Latvian, Mongol, Portuguese, Swahili and Thai). In its second phase (1999 – 2005) the project will seek to further extend UNL access.

The UNL system basically consists of a network and a conversion program between UNL and native, involved languages.

Conversion consists of enconverter and deconverter. Enconverter translates from the native languages to UNL; deconverter translates from UNL to the native language. All this activity is carried out via the network.
UNL Language Centers

Since the year 2000, it was decided by UNU/IAS to establish a language center for each of the 16 participant languages. The UNL Arabic Language Center (ALC) was established at The Royal Scientific Society (RSS) in Amman, Jordan. A description of the main activities of the UNL ALC is illustrated below.

How work is done at UNL ALC

The UNL ALC recruits a 7-member team headed by the Computer Technology, Training and Industrial Studies Center (CTTISC) Director at the Royal Scientific Society (RSS). These members are computer programmers, Analysts and a linguist. They work on the dictionary, the enconverter, the deconverter, the UNL Proxy, the Org-Explorer, the UNL Editor and other relevant softwares.

Dictionary

The Arabic UNL dictionary consists of 79000 words. It is made up of UWs, the Arabic equivalent, categories (conceptual), and semantic constraints. “Categories” means dividing words under main “categories”. UWs are divided into Nouns, verbs and particles. A noun may belong to one of the following categories: abstract, concrete, living, volitional or functional category. A verb may be transitive or intransitive, or it may be stative. In the first case it falls under (icl >do) e.g. He ate the apple; in the second case under (icl >occur) e.g. The car hit a wall; in the third case under (icl > be ) e.g. He is rich.

Semantic constraints of UWs means specifying their respective grammatical attributes, i.e. saying if the UW is noun or verb. If it is a noun the following constraints are added: singular / plural; count / non count; and definite or indefinite. If it is a verb, we add the following constraints, past / present; active / passive; affirmative / negative; and person (i.e. first, second or third person).

There are also “relations labels” which help narrow down the choice of a specific UW. Among these labels are “agt” which stands for the action initiator, “obj” for recipient of the action, “aoj” for the object
described or modified, etc. All adverbials are included under (icl > how).

The dictionary team have handled about 11000 words.

**Enconverter**

The enconverter is sort of software which changes or translates Arabic, for example, into UNL. To change an Arabic text into UNL, the text is primarily edited (by a machine editor) and afterwards handled or enconverted into UNL by means of the enconverter.

Example:

```
[<][{} "" (SHEAD) <.,0,0>;
[ ][{} "eat" (@entry,has_obj,has_sub,blk,&@present,Pr,M_infl,V,Tr,Rg,Aff,Act,3P) <A,0,1>;
[ ][{} <{V,has_sub,Tr,has_obj;+has_obj;}>[N::obj;](STAIL)P130;
[ >obj
[ [{} "eat" (has_sub,blk,&@present,Pr,M_infl,V,Tr,Rg,Aff,Act,3P) <A,0,1>;
[ ][{V,^has_sub:+has_sub;}{N::agt;}P130;
[ ]>agt
[ [ ]{} "eat" (blk,&@present,Pr,M_infl,V,Tr,Rg,Aff,Act,3P) <A,0,1>;
[ [ ]{} "sprout(ocl>young person)" (is_sub,blk,AL,&@def,Df,S,N,M,IrgPl) <A,0,1>;
[ [ ]{} "apple" (AL,&@def,Df,Ta_Added,LShift,S,N,IrgPl,Ind,FSx,F,~?) <A,0,1>;
[ ]{>>} "" (STAIL) <.,0,0>;
[S]

;obj(eat:01.@entry.@present, apple:0D.@def)
agt(eat:01.@entry.@present, sprout(ocl>young person):07.@def)
[/unl]
[/S]
;Time 0.0 Sec
;Done!
```

The enconverter team have completed about 300 rules.

**Deconverter**

The deconverter is a system to automatically convert UNL into other (involved) languages. The deconverter plays a core role in the UNL system as it generates the native language from UNL. It expresses UNL information with 100% accuracy. It follows that information once composed in UNL, can be translated to, and understandable in any language as long as there is a deconverter of the language. It can be developed by interfacing existing system with UNL.

Example
ITU works toward developing and managing the world’s telecommunication resources.
The deconverter team have completed about 1000 rules.

Work on the UNL began in 1996, although it was interrupted several times. Arabic language center (ALC) is the sole center in the Arab world in charge of handling Arabic Language which is one of the six UNL languages. It is in the Computer Technology, Training and Industrial Studies Center at the Royal Scientific Society (CTTISC) in Amman, Jordan. The RSS, among other things, conducts technical and scientific research.

The ALC recruits a 7-member team headed by the CTTISC Director. These members are computer programmers, analysts and a linguist. They work on the dictionary, the enconverter, the deconverter and on the relevant software.

Teams of work: the dictionary team, the enconverter team and the deconverter team.
**UNL Proxy Server**

The UNL Proxy Server is a stand-alone application developed by java programming language and works in a terminal computer. It works as a filter that allows Internet browsers to recognize web pages written in UNL and engages the appropriate Language Server on the Internet so that the document can be read in a natural language.

The UNL Proxy server works in the following way. First, the user must adjust his or her browser settings to use the UNL Proxy Server. He or she then starts the Proxy server and sets it to access a desired Language Server. In the process of accessing a web page, the browser will pass the URL through the UNL Proxy. The UNL Proxy determines whether the file has a UNL extension and contains UNL expressions. If this is confirmed, the Proxy communicates with the appropriate Language Server specified by the user. The Language Server then “deconverts” the UNL expressions into the desired natural language, the result of which is sent back by the UNL Proxy to the browser. In case only HTML files are found, the Proxy allows them to be read as they appear.

In this manner, if a user adjusts his or her browser settings to use the UNL Proxy server and selects Arabic language for the UNL Proxy server, he or she will then view a web page written in UNL in Arabic. Figure 8.5.2 shows the interface of settings of the UNL Proxy Server for selecting a language.

![Diagram of UNL Proxy Server Architecture](image-url)
1. Functionalities of the UNL Proxy Server

Figure 8.5.3 shows the interface and functionalities of the UNL Proxy Server. It receives the URL address of a web page, accesses the appropriate language sever if the web page has a UNL extension or a UNL document exists (will be explained later on)—includes UNL documents—and returns the web page in the required language to the browser for displaying.

The function of the UNL Proxy Server is divided into two levels. In the first level, the UNL Proxy Server functions as follows:

1) The user accesses a web page using a web browser by clicking or requesting its URL address. For instance, like “http://www.undl.org/unlweb.unl” or “http://www.undl.org/unlweb.html”
2) The UNL Proxy Server receives the URL and checks whether it is requesting a web page that includes UNL documents or not. This operation is carried out by checking whether the extension of the URL address is “UNL”.
3) If the URL is requesting an UNL web page, for instance like “http://www.undl.org/unlweb.unl”, the target web page with the UNL extension will be retrieved from the Internet and treated as including UNL documents to be passed to the process of “Language Server Access”.
4) If the URL is not requesting a UNL web page, the UNL Proxy Server will then check whether its extension is “htm” or “html”, and this and further processes are carried out in “Other Processing” as the second level of process, which will be explained below.
The UNL Proxy will finally send the result of either a web page that is deconverted from UNL or a non-UNL original web page to the browser for displaying.

Figure 8.5.4 “Other Processing”

In the second level, the UNL Proxy Server will check whether there is an UNL web page even if it is not requested explicitly, and carry out the relevant processes. For behaving so, the UNL Proxy Server need to be always active in the terminal computer and a UNL web page has to be developed using the same name as normal web page but with extension “unl”. Figure 8.5.4 shows data and process flow of Other Processing.

1. In this stage the UNL Proxy Server will first check if the requested web page is either “htm” or “html” by checking its extension.

2. If the type of the requested web page is neither “html” nor “htm”, the requested file, however, will be retrieved from the Internet and to be sent to the user as a response to her or his request.

3. If the type of the requested web page is either “html” or “htm”, the UNL Proxy will then change its extension to “unl” instead of “html” or “htm”. For instance, if “http://www.undl.org/unlweb.html” is requested, it will be changed to “http://www.undl.org/unlweb.unl”.

(5)
(4) The UNL Proxy will try to retrieve the UNL web page with the extension “unl”.
(5) If the UNL web page exists, it then is treated as including UNL documents and to be passed to the process of “Language Server Access” for deconversion.
(6) The web page with deconverted results will be sent to the user as a response for his request.
(7) If the UNL web page does not exist, the UNL Proxy will then retrieve the original web page using the URL with extension “htm” or “html” from the Internet and send it to the user as a response to his request.

![Diagram of Protocol of Language Server Access](image)

*Figure 8.5.5 Protocol of Language Server Access*

Figure 8.5.5 shows the protocol of Language Server Access. The details are the following:

1. The UNL Proxy Server first determines the Language Server according the language information selected by the user using the UNL Proxy GUI.
2. It then connects to the Language Server through a call to the Deco CGI and passes the user name, password and UNL documents to it.
3. The Language Server starts the deconversion of the UNL documents using the DeConverter. The UNL documents are to be deconverted into the required target language.
4. As soon as the deconversion is completed, the UNL Language Server returns the texts of target language to UNL Proxy through the DeCo CGI.
The Org Explorer

This is a program that enables the user to search and find information relating to registered companies and organisations by utilising the UNL Proxy Server. A demo of this Org Explorer was made in Geneva on the 3rd of July 2001. In this demo, only the, International Communication Union (ITU), was registered for exploration.

UNL Editor

This is an editor that enables converting a natural language text to UNL language. The resulting UNL excerpt is convertible back to the original natural language using the very same editor. This editor is useful in automatic conversion to, and from, UNL. The editor will be further enhanced to employ natural language-specific tags for pre-editing in order to generate more precise encoding to UNL.

UNL Activities:

- November 1999 a general conference about UNL was held at Brussels, Belgium.

- January 2001 a conference and presentations took place in Geneva Switzerland in which the ALC presented the UNL Proxy.

- May 2001 a workshop / Training for Dictionary Enconverter and Deconverter was held in Geneva, Switzerland.

- June 2001 a Developers meeting was held in Indonesia in which the ALC presented the Org-Explorer.

- July 2001 a conference / Presentations of UNL projects, Org-Explorer was held in Geneva, Switzerland.

- November 2001 a conference on building global knowledge was held in Shangahi, China.
The future of UNL: Multi-Language Information Platform

By way of conclusion the:

- UNL system is positioned as one of the components of the information infrastructure that can smooth international exchange of information. Information could be transmitted all over the world from international organizations, the information industry, and individuals in offices and homes, without having to consider language barriers. Information could also be received from anywhere in the world.

- A multi-language information platform that used UNL would not only enable communication using a user's mother tongue to many other languages. It would also remove the language barriers in every area of the information technology world. If information from different language was converted into UNL in various software treating language, and in information resources such as information bases and data bases, the software and information resources could be shared throughout the world.